Supplementary Information

Online supplement for

Real time prediction of size resolved ultrafine PM on freeways

Srijan Aggarwal, Ricky Jain, Julian D. Marshall

This supplement provides a more detailed description of the stratification method, a summary of parameters that were investigated in the analyses, then the following tables and figures. Table S1 is an expanded version of Table 1. Table S2 shows the stratification scheme used in this work. Table S3 shows speeds, traffic volumes and particle count information for the strata included in the study. Table S4 provides summary statistics for the input variables (weather; traffic) used in this study. An expanded list of stepwise regression models sub-divided by wind-direction and measurement year are presented in Table S5 and S6, respectively. Table S7 presents the results from bootstrap analyses.

Figure S1 shows four color-coded maps for measured and predicted (based on the mean and median predicted models) PNC along the test route for a single day and for ten days respectively (to contrast short-term versus long term prediction results). Figure S2 shows plots for correlations between predicted and observed PNC values for 10^{th} percentile, 25^{th} percentile, median, 75^{th} percentile, 90^{th} percentile, mean and standard deviation. Figure S3 shows correlations for observed PNC values with average loop speed and traffic loop volume. Figure S4 shows color coded maps for average loop speeds and total loop volumes along the test route, averaged for the same ten days as in Figure S1. Figure S5 shows the non-normalized coefficients of weather parameters for size resolved predictive models.

Stratification Method

Strata were selected so as to achieve even distribution of data among the strata. To ensure that our results are not contingent on the stratification scheme, we experimented with systematically varying the strata as a sensitivity analysis; twenty different stratification schemes were evaluated. For each alternative stratification scheme, we generated regression models. Speed and volume coefficients in the models exhibited low coefficients of variability, and model predictions were highly correlated with each other. Thus, we did not find evidence that model results are highly sensitive to stratification scheme. The most efficient stratification scheme was selected based on three criteria: model performance (measured by adjusted R^2), statistical significance of the model coefficients, and the percentage of minute-averaged data set used. The optimal stratification scheme selected had forty-two strata (six categories for traffic volumes and seven categories for traffic speeds) and used 96% of the valid data obtained. For each strata, we calculate mean values for traffic parameters (speed, volume); weather parameters (temperature, wind speed and relative humidity [RH]); and, for the dependent variable (log-transformed particle counts), several percentiles (10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles), mean and standard deviation. Also, median of log-transformed particles counts were determined for each of the 32 EEPS channels. Example measurements are shown in Table S3 for total PNC.

Summary of Parameters Investigated

(1) **Particle size.** We modeled total particle number concentration (all bins combined), and then also separately modeled PNC in each size-bin. Figures 4 and 6 illustrate results by size-bin.

(2) **Traffic speed and volume.** These are the core independent variables; a specific combination of traffic speed + volume is one "strata". Strata are listed in Table S3.

(3) **Concentration percentiles (P10, P25, P50, P75, P90, mean, st dev), for total particle number concentration.** In general, for a given strata, real-time concentrations will vary over time. This short-term variability is driven by small-scale turbulence: one second, the concentration is low (air happens to be cleaner in that one second), the next second the concentration is high (air happens to be dirties right then). These temporal variations in concentration happen even for a fixed set of conditions (speed = "x", volume = "y"). Thus, the models predict several percentiles, e.g., P10 is a model that predicts *cleaner-air* measurements during specific conditions (approximately, "on-road background" for those conditions), whereas P90 is a model that predicts *dirtier-air* measurements during specific conditions (approximately, "on-road in-plume"). The lower plot for Figure 5 shows these percentiles of concentration.

(4) **Weather.** Models were developed that included and excluded weather (wind speed, temperature and RH) as independent variables.

Table S1: Summary of studies on real time on-road measurement of ultrafine particles (UFPs).

Reference	Sampling year	Location	Methods	Pollutants	Instruments	Number of hours	Key results
Bukowiecki 2002a [17]	1999-2000	Minneapolis, Minnesota. Columbus, Indiana	Average surface diameter and PAS/DC versus Avg surface diameter scatter plots.	UFP	photoelectric aerosol sensor(PAS), diffusion charging sensor (DC), CPC, SMPS.		Theoretical conclusions related to aerosols.
Bukowiecki 2002b [13]	2001	Zürich (Switzerland)	Design and construction of mobile laboratory, diurnal variation of aerosols.	UFP	CPC*	4 hours	1. This case study confirms that there is a large diurnal and regional variation of ultrafine particles for both urban and rural areas. 2. Neither the UFP (<50 nm) nor the total particle number concentration is an exclusive indicator of primary traffic emissions. Diurnal, regional and other dynamic variability factors are also involved.
Canagratna 2004 [15]	2000-2001	New York city, New York.	Vehicle chase studies, chemical composition, size distribution	UFP, PM, NO, NO2, CO, N2O, CH4, SO2, and HCHO.	Aerodyne, aerosol mass spectrometer (AMS), CPC, Tunable Diode Lasers (TDL), LiCOR	~24 hours	1. The nonrefractory diesel exhaust PM appears to be dominated by lubricating oil and the typical measured mass distribution of organic as well as sulfate species. 2. Order of PM emissions in diesel operated engines : 6V-92 engines> series-50 engines > CNG engines
Weijers et al. 2004 [12]	1999-2000	Amsterdam, Netherlands	Dispersion study, particle size distributions, regional variability, within city variability	UFP, PM	CPC and an optical aerosol spectrometer	3 days	1. Aerosol concentrations decrease exponentially with increasing distance from the road. 2. Number concentrations are more sensitive than mass concentrations due to dominance of number of UFPs. 3. Number concentration in city change on scale of a hundred meters, correlating with the local traffic intensity and driving conditions

Kittleson et al. 2004 [11]	2000	Minneapolis, Minnesota	Size distributions, fuel-specific emissions, dispersion study.	UFP, CO, CO_2, NO_x.	CPC, SMPS, IR CO analyzer, IR CO_2 analyzer, chemiluminescence NOx analyzer.	~ 20 hours	1. High UFPs correlated with high speed traffic. 2. Most of the particles added by the on-road fleet were below 50nm in diameter. 3. Number concentrations measured in residential areas, 10–20m from the highway, considerably lower than on-road concentrations, and much lower concentrations for areas 500-700 m from the highway.
Westerdahl et al. 2005 [8]	2003	Los Angeles, CA	Pollutant concentration differences by location, pollutant concentration correlations, UFP size distributions, time series plots	BC, NO, PM-PAH,UFP, NO2, CO, CO2, $PM_{2.5}$	CPC, SMPS, Aethalometer, DMA, PAH Analyzer, NOx analyzer, Q-trak plus monitor, TSI DusTrak	12 hours	1. Good correlation between UFP concentrations and BC, NO and PM-PAHs. 2. Freeway concentrations an order of magnitude higher than on residential streets for UFP, NO, BC and CO. 3. Average concentrations of UFP and related pollutants varied strongly by location, road type, and truck traffic volumes, suggesting a relationship between these concentrations and truck traffic density.
Pirjola et al. 2006 [10]	2003-04	Helsinki, Finland	Seasonal (summer vs winter analysis).Pollutant dispersion measurements upto 140 m distance from the road side. Particle size distribution analysis in the range of 3 nm -10 um.	CO, NO, UFP, NO_x,PM	CPC, SMPS, CO and NOx monitors	384 hours	1. Average concentrations 2-3 times higher in winter than in summer. 2. Concentrations fell to 20-40% as far as only 65 m away from the road, still more than double the background urban concentrations. 3. 85% or more particles were smaller than 50 nm. Observed distribution was multi-modal.
Fruin et al. 2008 [17]	2003	Los Angeles, CA	DVD analysis, multiple regression, ANOVA	BC, NO, PM-PAH,UFP, NO2, CO, CO2, $PM_{2.5}$	CPC, SMPS, Aethalometer, DMA, PAH Analyzer, NOx analyzer, Q-trak plus monitor, TSI dustrak	15 hours	1. Arterial concentrations one-third to freeway concentrations. 2. Freeways responsible for 33-45% UFP exposure in LA 3. Diesel powered vehicles primary sources of UFPs, NO, BC, PM-PAH

Johnson et al. 2009 [21]	2006-2007	Minneapolis, MN	Fuel specific apportionment of particle number concentration on highways, size distribution statistics	UFP, CO, CO2, NOx	CPC, SMPS, EEPS, TSI DustTrak, NDIR QTrak	85 hours	Decrease in fuel sulfur content led to reduced particle counts from 2002 to 2007.
Int Panis et al. 2010 [13]	2009	Belgium	Comparison of fine particle exposure to car riders and cyclists in three Belgian cities	UFP, $PM_{2.5}$, PM_{10}	TSI P-Trak, TSI DustTrak	~30 hours	Mean bicycle/car ratio for PNC and PM are close to 1 and rarely significant.

[*] Other instruments were used –e.g., to measure gaseous pollutants – but the manuscript cited focused on instrument(s) listed here.

Table S2: Two-way stratification scheme, showing strata serial numbers (see Table S3 for data per serial number), based on average loop speed and total loop volume

Total loop volume (veh min^{-1})	Average loop speed (km h^{-1})						
	24 to 64	64 to 80	80 to 87	87 to 93	93 to 100	100 to 105	> 105
< 40	1	7	13	19	25	31	37
40-50	2	8	14	20	26	32	38
50-60	3	9	15	21	27	33	39
60-70	4	10	16	22	28	34	40
70-80	5	11	17	23	29	35	41
>80	6	12	18	24	30	36	42

Table S3. Speed, traffic volumes and particle count information for the strata included in the study

Strata serial number[a]	Number of data points	Average loop speed[b] (km h^{-1})	Total loop volume (veh min^{-1})	PNC [c] (Median Log)
1	20	53	35	4.3
2	25	53	45	4.4
9	31	76	55	4.45
10	52	76	65	4.55
11	37	76	75	4.57
12	22	76	88	4.82
14	44	84	46	4.41
15	84	84	56	4.58
16	142	84	65	4.64
17	81	84	74	4.68
18	21	84	86	4.66
20	70	90	46	4.38
21	213	90	55	4.58
22	258	90	65	4.63
23	110	90	74	4.61
24	29	90	86	4.50
25	21	96	36	4.11
26	84	96	46	4.29
27	181	96	56	4.47
28	175	96	65	4.53
29	53	96	74	4.59
32	42	102	46	4.27
33	77	102	55	4.26
34	52	102	65	4.52
37	20	110	33	4.07
38	54	110	45	4.15
39	81	110	55	4.26
40	43	110	63	4.4

[a] See Table S2. For example, Strata #1 (first row) refers to times when the loop volume is <40 veh min^{-1} and traffic speed is 24-64 km h^{-1}. Only cells with greater than 20 data points were included in the analysis. [b] Ensemble averages for the cells with same speed range were calculated, because individual cell averages for the same speed range were similar. [c] Values are the base-10 logarithm of the median. For example, in the first row log median is 4.30, indicating a particle concentration of 19,900 particles cm^{-3}.

Table S4: Summary statistics for the independent variables

Statistical parameter	Temperature (deg C)	RH	Wind speed (km h^{-1})	Traffic speed (km h^{-1})	Traffic volume (veh min^{-1})
Mean	25	0.46	16.0	91.3	60.1
Median	25	0.44	14.8	91.6	60.1
Std Dev	3.6	0.11	8.3	11.3	11.7
SE	0.08	0	0.18	0.24	0.25
CV (%)	14%	23%	52%	12%	19%
Minimum	17.8	0.28	0	36.7	23.4
10^{th} percentile	20.6	0.34	5.6	80.3	44.9
25^{th} percentile	22.2	0.37	9.3	86.3	52.3
75^{th} percentile	27.8	0.55	20.4	97.4	67.4
90^{th} percentile	30	0.61	27.9	104	74
Maximum	33.9	0.8	37.0	130	101

Table S5: Multi-linear regression models for prediction of particle number concentration (PNC) based on wind speed classification.

Dependent variable[a]	Condition[e]	Un-aided stepwise regression[b]			Aided stepwise regression[c]		
		Parameter[d]	**R^2**	**adj-R^2**	**Parameter**	**R^2**	**adj-R^2**
Median	**upwind**	**RH**	0.66	0.64	**RH**, WS	0.70	0.68
	downwind	Vol, Temp, **RH**, WS	0.73	0.69	Same as un-aided model		
	neither	**RH**	0.72	0.71	Same as un-aided model		
	total	**RH**, WS	0.78	0.76	Speed, Vol, Temp, **RH**	0.85	0.83
					Vol, Temp, **RH**	0.83	0.81
					Vol, **RH**, WS	0.81	0.78
					Speed, Vol, **RH**, WS	0.83	0.81
					Speed, Vol	0.79	0.77
Mean	**upwind**	**Vol**, RH, WS	0.76	0.73	Speed, **Vol**, RH, WS	0.79	0.76
					Vol, RH, WS	0.76	0.73
	downwind	**Vol**	0.46	0.44	**Vol**, Temp, WS	0.66	0.62
	neither	**Vol**, RH	0.73	0.70	Same as un-aided model		
	total	Speed, **Vol**	0.79	0.77	Speed, **Vol**, Temp	0.81	0.79

					Vol, Temp, RH	0.82	0.80
10th Percentile	**upwind**	Speed, Vol, WS	0.73	0.69	Same as un-aided model		
	downwind	Vol, Temp	0.41	0.36	Same as un-aided model		
	neither	RH	0.36	0.34	Speed, Vol	0.53	0.49
	total	Speed, Vol	0.60	0.57	Same as un-aided model		
25th Percentile	**upwind**	Speed, Vol, WS	0.75	0.72	Vol, RH, WS	0.69	0.65
					Vol, RH	0.64	0.61
	downwind	Vol, Temp	0.60	0.57	Same as un-aided model		
	neither	RH, WS	0.76	0.74	Speed, Vol	0.78	0.76
	total	Speed, Vol, Temp	0.84	0.82	Same as un-aided model		
75th Percentile	**upwind**	**RH**	0.55	0.54	**RH**, WS	0.61	0.58
	downwind	**RH**	0.46	0.44	Temp, **RH**, WS	0.64	0.59
	neither	**RH**	0.63	0.62	Speed, **RH**, WS	0.72	0.68
	total	Vol, **RH**	0.77	0.75	Vol, Temp, **RH**	0.80	0.77
90th Percentile	**upwind**	**RH**, WS	0.65	0.62	Vol, **RH**, WS	0.69	0.65
	downwind	**RH**	0.31	0.29	Temp, **RH**, WS	0.54	0.48
					Speed, **RH**	0.38	0.34
	neither	Vol	0.46	0.44	**RH**	0.40	0.38
					Vol, Temp, **RH**	0.74	0.71
	total	Vol	0.66	0.64	Speed, Vol, Temp, **RH**	0.78	0.74
					RH	0.54	0.53
Standard Deviation	**upwind**	None			Speed, Temp, RH	0.43	0.36
	downwind	WS	0.17	0.14	Vol, Temp, WS	0.38	0.30
					Vol, Temp	0.28	0.22
					Vol, WS	0.28	0.22
	neither	Speed	0.41	0.39	Speed, Vol, RH, WS	0.64	0.58
	total	Speed	0.46	0.44	Speed, Vol, RH	0.66	0.62

[a] Dependent variable: Particle number concentrations (PNC; units: cm^{-3}); [b] Un-aided stepwise regression refers to forward stepwise regression starting with a null model. [c] Aided stepwise regression refers to a combination of forward and backward (selective removal starting with a full model) stepwise approach. [d] Independent variables: strata average of loop speed (Speed; units: km h^{-1}); total traffic volume (Vol; number of vehicles per minute); urban temperature (TEMP; °C); relative humidity (RH); and, wind speed (WS; km h^{-1}). All coefficients have $p<0.10$. Within each strata of traffic speed and volume**,** 25^{th} percentile, median, mean and standard deviation were computed for log transformed data set for PNC. [e] Wind direction condition.

Table S6: Multi-linear regression models for prediction of particle number concentration (PNC) based on different measurement years.

	median log(PNC) model[a,c]		mean log(PNC) model[a,c]	
	Year 2006	Year 2007	Year 2006	Year 2007
constant term	4.20	4.22	4.15	4.26
SP coefficient[b]	-0.009	-0.002	-0.006	-0.003
VOL coefficient[b]	0.014	0.006	0.012	0.005
Adj-R^2	0.73	0.50	0.71	0.50

[a] Dependent variable: Particle number concentrations (PNC; units: cm^{-3}); [b] Independent variables: strata average of loop speed (SP; units: km h^{-1}); total traffic volume (VOL; number of vehicles per minute); **[c]** Within each strata of traffic speed and volume**,** median and mean were computed for log transformed data set for PNC.

Table S7: Bootstrap re-sampling results for the particle count dataset used in this study

	Original dataset		Boostrap datasets				
	Values	SE	min	mean	max	SDOM	CV
25th percentile log PNC							
Constant	1.373	1.324	-1.082	3.074	7.991	1.303	42%
Speed coeff	-0.011	0.002	-0.017	-0.011	-0.002	0.002	20%
Volume coeff	0.010	0.001	0.005	0.010	0.013	0.001	14%
Temp coeff	0.037	0.017	-0.054	0.015	0.069	0.017	111%
Adj R^2	0.823		0.285	0.699	0.885	0.096	14%
50th percentile log PNC							
Constant	4.237	0.128	3.803	4.221	4.605	0.116	3%
Speed coeff	-0.007	0.002	-0.011	-0.006	0.000	0.002	29%
Volume coeff	0.010	0.001	0.006	0.010	0.014	0.001	12%
Adj R^2	0.771		0.424	0.685	0.878	0.075	11%
Mean log PNC							
Constant	4.175	0.118	3.895	4.178	4.474	0.094	2%
Speed coeff	-0.005	0.002	-0.009	-0.005	-0.001	0.001	28%
Volume coeff	0.009	0.001	0.006	0.009	0.012	0.001	9%
Adj R^2	0.769		0.454	0.698	0.858	0.067	10%
SD log PNC							
Constant	0.132	0.053	-0.075	0.115	0.351	0.076	66%
Speed coeff	0.004	0.001	0.001	0.005	0.008	0.001	29%
Adj R^2	0.442		-0.033	0.316	0.719	0.159	50%

[a] Bootstrap sample size: n=1000.[b] SE is standard error. [c] SD is standard deviation of mean. [d] CV is the coefficient of variation.

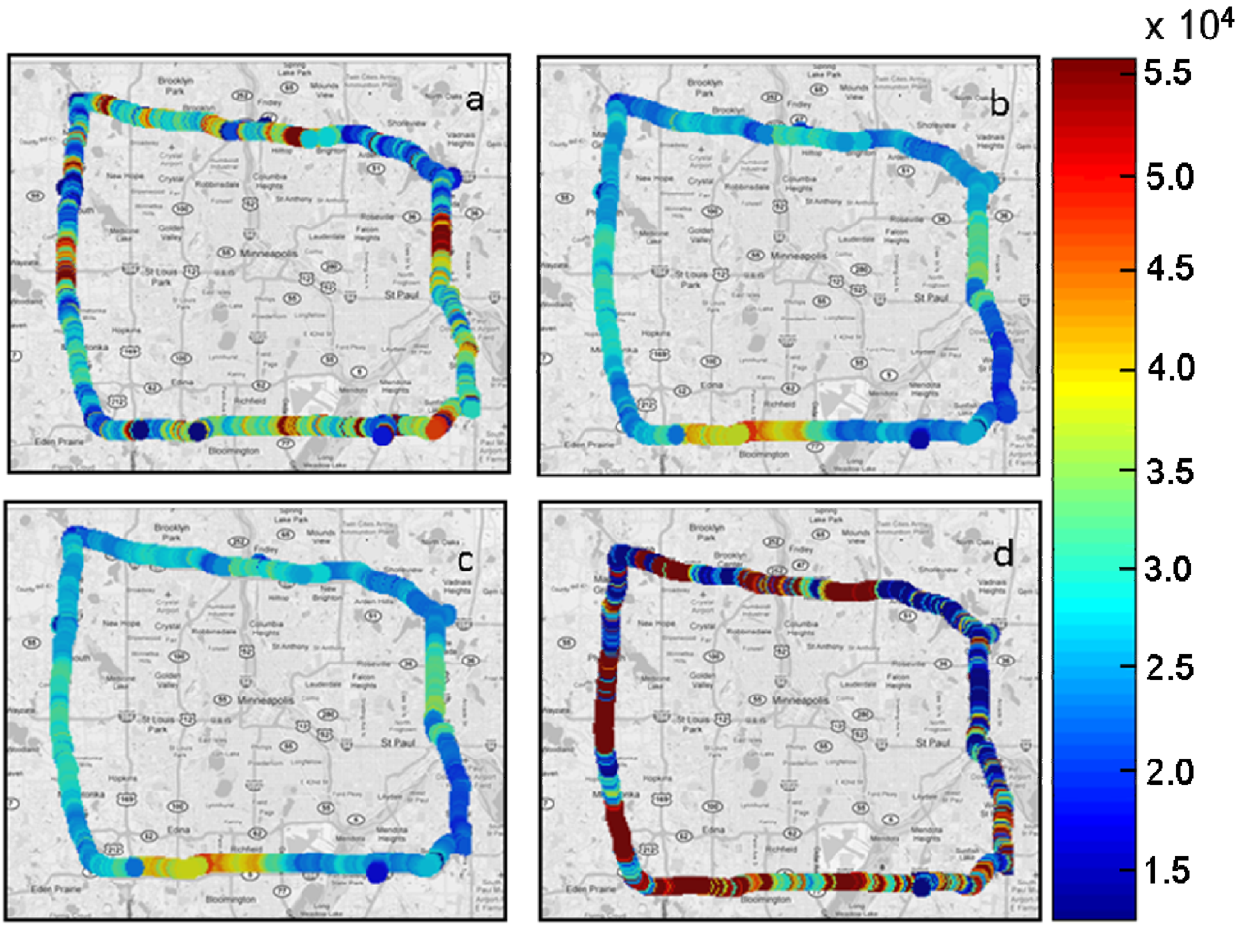

Figure S1: Particle number concentrations along the test route (a) measured concentrations averaged over ten summer days in June 2007 (b) median predicted concentrations averaged over ten summer days in June 2007 (c) mean predicted concentrations averaged over ten summer days in June 2007 (d) median predicted concentrations along the test route on a representative day (28th June, 2007; 9:00 AM to 1:30 PM). MNDOT-measured vehicle speeds and volumes were used to make predictions using the proposed models.

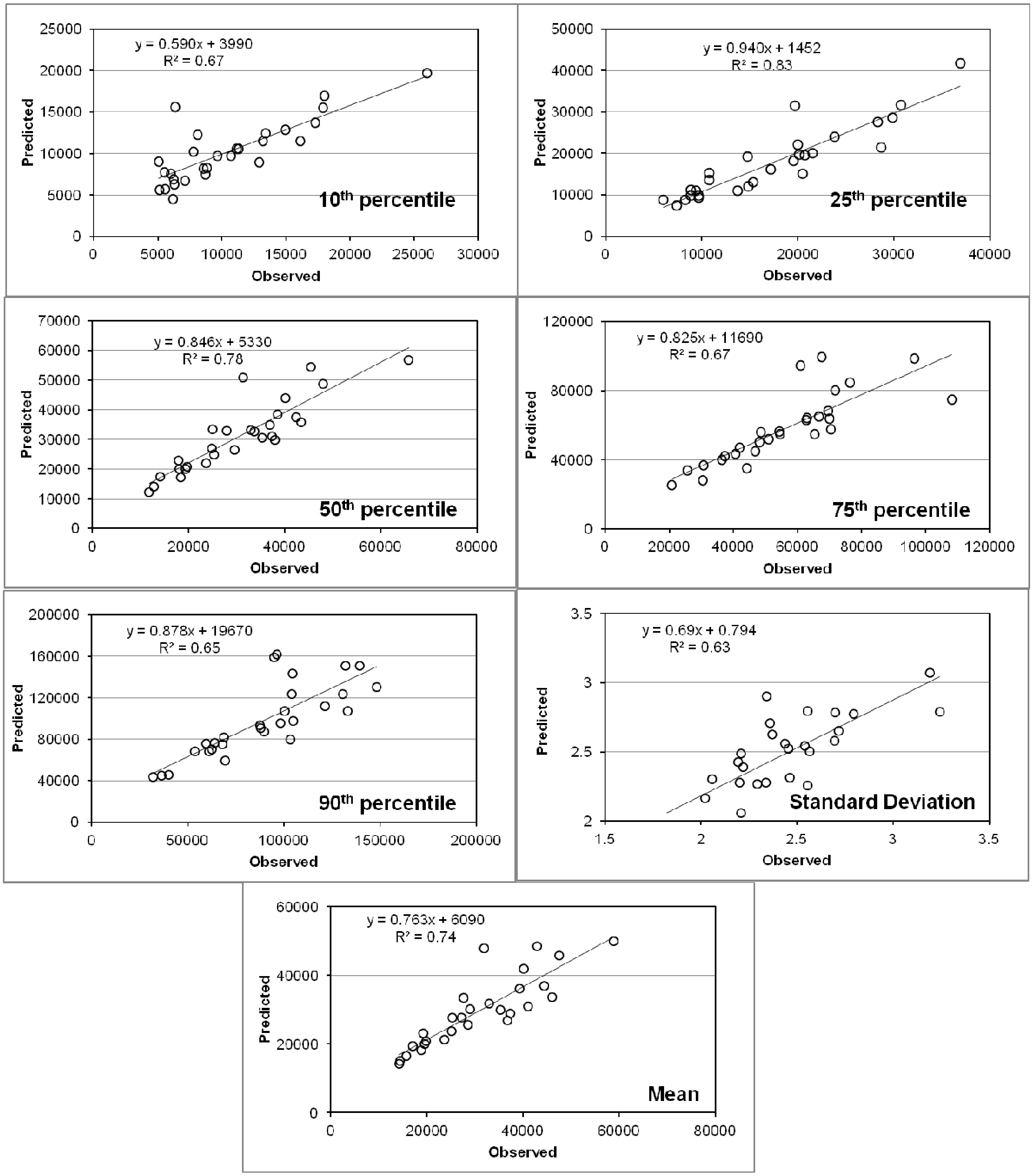

Figure S2: Regression plots for predicted versus observed PNC values for each of 28 strata (as shown in Table S3), for 10th, 25th, 50th,75th, 90th percentiles; mean and standard deviation. Note that the R^2 values presented here are different from those presented in Table 2/Figure 3 of the main article; as the R^2 values reported therein represent model performance derived from regression analyses of the particular models.

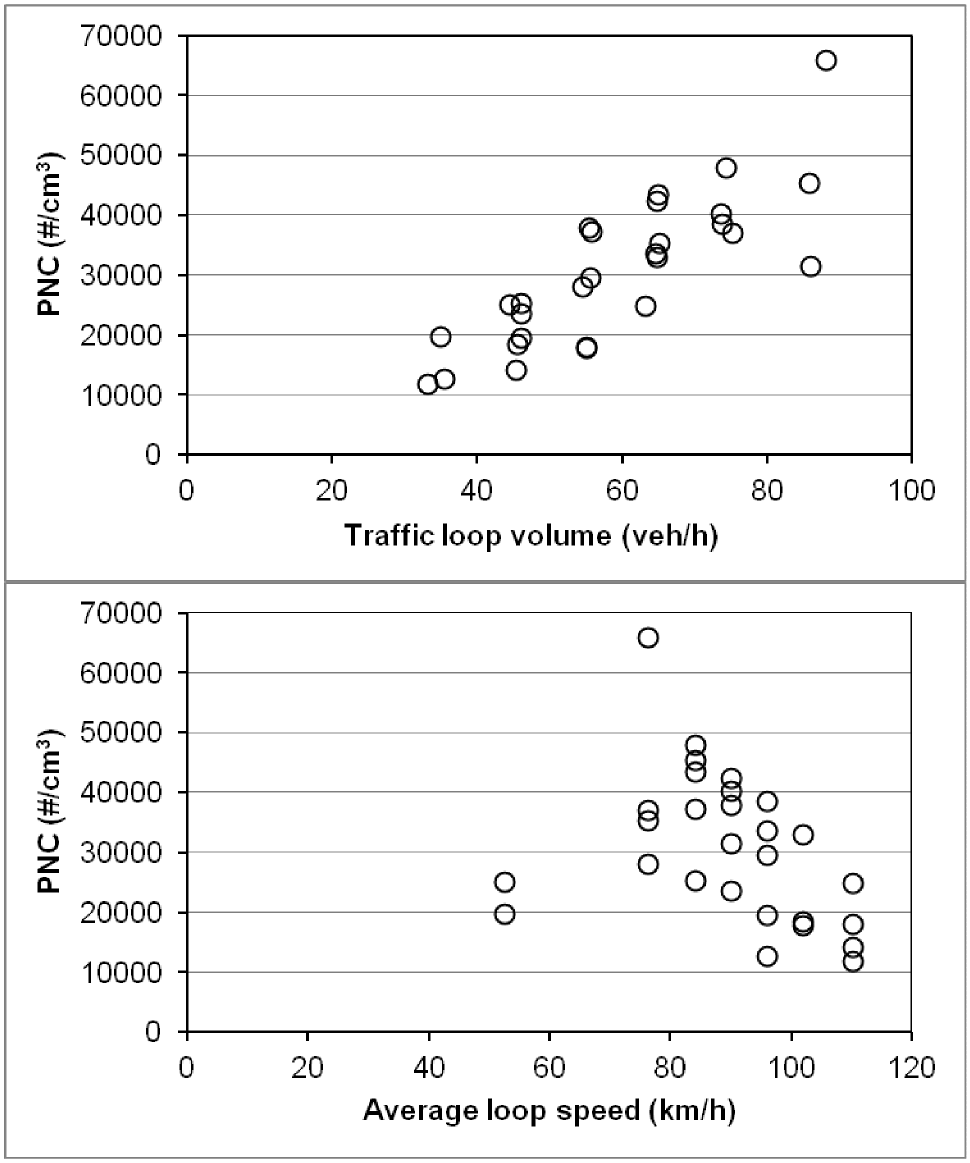

Figure S3: Scatter plots for observed PNC values with average loop speed and traffic loop volume for each of the 28 strata (as shown in Table S3).

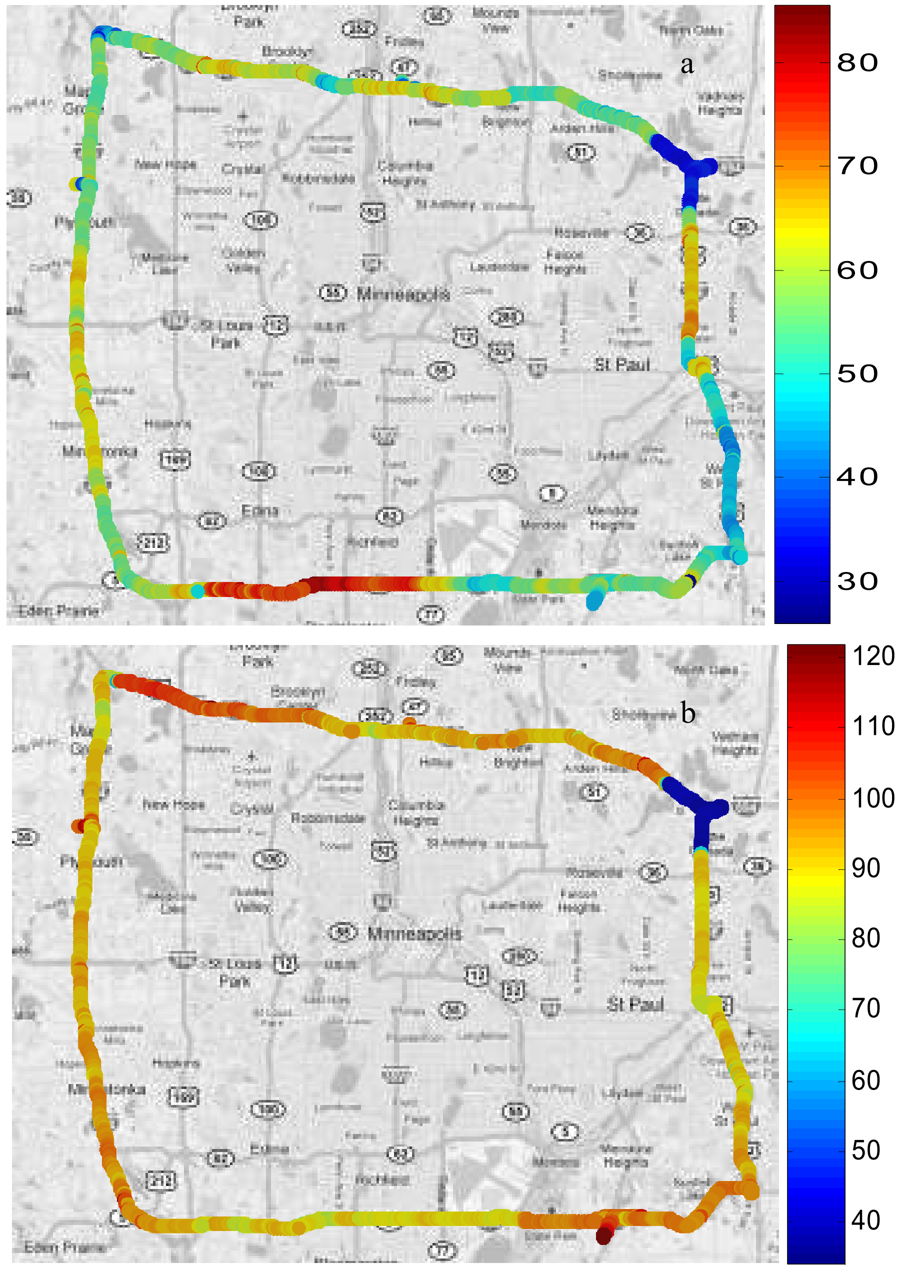

Figure S4: Measured values of (a) total traffic volumes, and (b) mean loop speeds, averaged over ten summer days in June 2007. These speeds and volumes were used to make prediction plots in Figure S1.

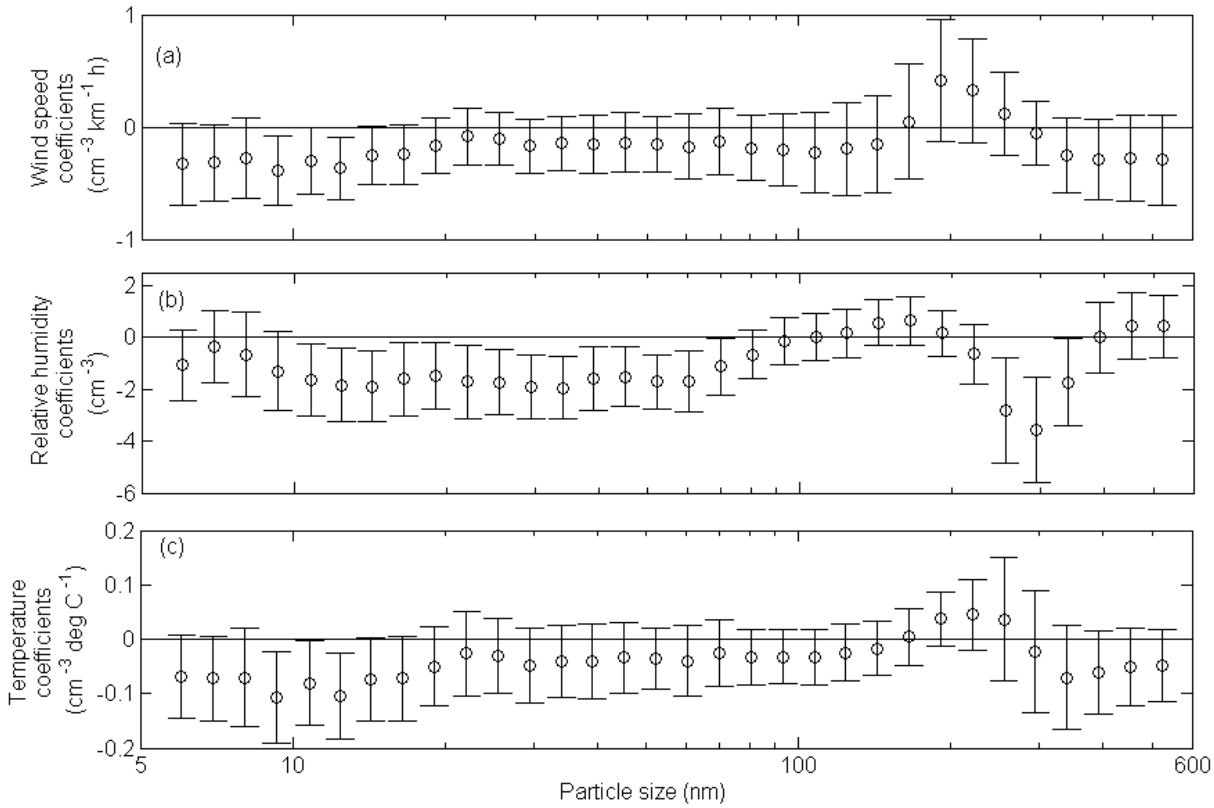

Figure S5: Coefficients (non-normalized) of meteorological parameters for size resolved prediction of median particle count measurements, from the 'inclusive weather' models. (a) temperature coefficients, (b) relative humidity coefficients, and (c) wind speed coefficients for the thirty-two models. The error bars represent 95% confidence intervals.